

**PSI RESEARCH DISCUSSION PAPER 1**

**Evaluating Training Programmes for the  
Long-Term Unemployed**

The study described in this paper was funded by the Department for Education and Employment, and was carried out jointly with Clive Payne and Martin Range of Oxford University and Steve Lissenburgh of the Policy Studies Institute. The interview surveys were designed and conducted by the National Centre for Social Research, where Kerry Sproston took main responsibility for the work.

**PSI RESEARCH DISCUSSION PAPER 1**

**Evaluating Training Programmes for the  
Long-Term Unemployed**  
An illustration of the matched comparison group  
methodology

Joan Payne



---

Policy Studies Institute

© Policy Studies Institute, 2000

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic or otherwise, without the prior permission of the copyright holder(s).

ISBN 0 85374 777 6

PSI Report No 871



Policy Studies Institute

For further information contact

Publications Dept, PSI, 100 Park Village East, London NW1 3SR

Tel (020) 7468 0468 Fax (020) 7468 2211 Email [pubs@psi.org.uk](mailto:pubs@psi.org.uk)

*PSI is a wholly owned subsidiary of the University of Westminster*

## **Abstract**

This paper describes an evaluation of a skills training programme for long-term unemployed adults in the UK, for which a random allocation design had been ruled out. After an initial survey of programme leavers, a matched comparison group was constructed from administrative records, based on claim dates, sex, age and locality. The close match achieved was partially undermined by subsequent sample attrition, necessitating a re-matching process. Results showed a high degree of selection into the Training for Work Programme even after matching. Subsequent improvements in administrative data in the UK have increased the viability of this type of design, and where analysis can be carried out entirely with administrative data, the problem of sample attrition is overcome.

### **Keywords**

Evaluation; matched comparison group; skills training; long-term unemployment; employment outcomes.



## Introduction

To measure the effect of a social programme on participants, we must answer the counter-factual question of what would have happened to them without the programme. Usually this cannot be done by simply comparing outcomes for participants and eligible non-participants, as the former are rarely a random sub-sample of the group at which the programme is aimed. On the contrary, they are usually selected (by themselves or others) on the basis of characteristics that may raise or lower their chances of securing the outcomes that the programme is designed to produce. Thus, for example, one skills training programme for the unemployed may attract people who are keen to find a job and the programme provider may admit only those who are likely to succeed, while the intake of another may comprise mainly people who have failed to find a place elsewhere or whose participation is in some way coerced.

In North America, random allocation research designs or ‘social experiments’ are often used to eliminate such selection biases in evaluation studies. Although there has been much debate about their merits, in many contexts social experiments offer considerable methodological advantages over alternative evaluation designs (Burtless and Orr, 1986; Lalonde, 1986; Manski and Garfinkel, 1992). In the UK, however, the government has only rarely permitted random allocation for evaluating social (rather than medical) programmes, the very few exceptions including, for example, the evaluation of Restart interviews for unemployed people carried out by White and Lakey (1992). The reasons for this reluctance have generally been ethical, political and pragmatic rather than methodological. Although arguments around these issues urgently need review, space prohibits such discussion here. However the practical impact has been to compel researchers to adopt alternative evaluation strategies.

This paper illustrates one such alternative, the matched comparison group design where the comparison group is constructed by purposive selection of individuals from administrative records. This strategy was used in a recent evaluation of a skills training programme for long-term unemployed people

in the UK, namely Training for Work (TfW) (Payne, Payne, Lissenburgh and Range, 1999).

The matched comparison group evaluation design is by no means unproblematic, as analysis results can vary with the method of constructing the comparison group and with the analytic model used (Fracker and Maynard, 1987). Nevertheless it generally offers a better chance of detecting small programme impacts than does comparison with a random sample of non-participants, because it reduces between-sample variance on observed confounding factors. However the extent of any remaining unobserved differences between the participant sample and matched comparison sample cannot be determined, and this can present difficulties when interpreting results (Firth, Payne and Payne, 1999). It follows that the lower the degree of selection of participants into the programme, the more likely it is that the matched comparison group design will give valid results.

One technique for constructing a matched comparison group that has several important advantages is propensity score matching (Lechner 2001, Heckman, LaLonde and Smith, 1999). At its simplest, this technique involves developing a statistical model for programme participation, and then constructing the comparison group from those non-participants with the highest fitted probability ('propensity') of participation. However, for propensity score matching to be carried out successfully, we need to have available enough data on both participants and non-participants to construct a good model of programme participation. This condition is met when there are good administrative data for non-participants, or when census data or existing large-scale surveys can be used. In the absence of such existing data sources, it could be viable to conduct a special-purpose survey of a random sample of non-participants to collect the data needed for propensity score matching, provided that we are confident in advance that participants are not highly selected into the programme. However if participants are highly selected, then it is unlikely that a random sample of non-participants (unless very large indeed) would throw up enough people with high propensity scores to enable us to construct a large enough comparison group (the 'support' problem). The method of constructing a matched comparison group described in this paper should be seen as an alternative to propensity score matching, to be considered when data collection is expensive and resources limit sample size.

The main part of this paper describes how a matched comparison group for a sample of TfW participants was constructed from administrative data, and assesses the closeness of the match that was achieved. A very brief outline of the evaluation results is given. The paper concludes with a discussion of the difficulties involved in carrying out this type of evaluation research, showing how recent advances in the availability of administrative data in the

UK can improve the design of evaluation studies. We begin, however, with a brief description of the TfW programme.

## TRAINING FOR WORK

The Department for Education and Employment (DfEE) launched Training for Work (TfW) in 1993, uniting features from a range of programmes developed over the previous two decades. Its aim was to help unemployed people find jobs and improve their skills, by providing appropriate training and work experience. Although TfW no longer exists as a separate programme, many of its elements survive as options under the various New Deals and in its direct successor, known as Work-Based Learning for Adults. TfW was open to men and women aged 18–60 who had been claimant unemployed for six months or more, although the latter condition was waived for certain groups with special needs, such as people with disabilities or with literacy or numeracy problems, labour market returners and former prisoners. During the year 1995/96, when the people involved in the evaluation study described here were on the programme, 235,900 people started training on TfW, and the planned budget for that year was £578 million (DfEE, 1997). Training providers were funded, via local Training and Enterprise Councils, according to a ‘starts and outcomes’ formula. This involved an initial payment for each entrant, with remaining payments contingent on whether the trainee achieved a ‘positive outcome’, which meant securing a job or obtaining a recognised qualification. The length of time that the trainee spent on TfW played no part in the funding formula.

Within TfW there were three main routes: employer placements, where the trainee gained work experience through working with an employer in an ordinary working environment; project placements, where the trainee worked with a voluntary organisation or joined a group to do work of benefit to the community; and full-time off-the-job training. Employer placements could be either with employed status, when the employer paid the trainee’s wages but received a subsidy for training costs, or with trainee status, when the employer received a training subsidy but did not pay wages (the trainee stayed on social security benefits). Those on employer or project placements often received off-the-job training as well, usually in a private training centre or a Further Education College. The mean stay on the programme of people who left TfW in autumn 1995 was around four months. Around two thirds had undertaken study for qualifications, almost all of which were vocational (Sproston and Smith, 1999).

## STUDY DESIGN

The aim of the evaluation study of Training for Work was to estimate the impact of the programme on participants' chances of getting a job, on the type of job that they got, and on their hourly pay. It began with a survey of a random sample of people leaving TfW in England and Wales in September or October 1995. Undoubtedly it would have been better to start with a survey of entrants rather than leavers, but the timetable of the sponsor (the DfEE) did not permit this. The sample was clustered in 80 'sampling points' (groups of adjacent postcode districts), and face-to-face interviews in March and April 1996 collected information on experiences on TfW, unemployment before entry, activities since leaving, and a range of personal characteristics. Full details can be found in Sproston (1999). In this Wave 1 survey, 1,498 interviews were achieved, representing 66 per cent of the issued sample. The biggest source of non-response – 19 per cent of the issued sample – involved problems with the addresses held on administrative records. A further 9 per cent of the issued sample claimed that they had never been on TfW, and not surprisingly, these included a disproportionate number of those with employed status and of those with very short stays on the programme. A Wave 2 survey of Wave 1 respondents who agreed to be re-interviewed took place in June and July 1997, when 1,060 interviews were achieved, representing 75 per cent of Wave 1 respondents and 43 per cent of the original issued sample. This wave updated information on activities since leaving TfW and collected work histories back to 1985.

The database used to select the matched comparison sample was the JUVOS (Joint Unemployment and Vacancies Operating System) Cohort. This held the start and end dates of all claims for unemployment-related benefits together with the sex, date of birth and postcode for, at that time, a 5 per cent sample of National Insurance numbers (NINOs). However, because it covered only 5 per cent of the TFW sample, JUVOS could not be used to provide matching information for TfW participants. Instead, variables equivalent to the JUVOS variables were derived from the Wave 1 interview data. Today JUVOS covers 100 per cent of NINOs plus some useful additional variables, and the matching process could be carried out entirely with JUVOS data, rather than from a mixture of survey and administrative data.

Obviously the variables used for matching had to be restricted to those available on JUVOS. The variable we gave most importance to was the length of the spell of unemployment that preceded entry to TfW, henceforth called the 'qualifying spell'. This was because the probability of leaving unemployment falls as the length of the unemployment spell increases, and because duration of unemployment can act as a proxy for other factors that affect a person's chances

of getting a job but are hard to measure, such as interview manner, motivation or personal appearance. The matching process involved the following steps.

1. Key characteristics of TfW participants were established from the Wave 1 survey interviews.<sup>1</sup> Respondents were excluded if they did not wish to be interviewed a second time, failed to give precise dates, or had entered TfW without being claimant unemployed for six months or more (the 'special needs' group). This left 1,038 out of 1,498 Wave 1 respondents available for matching.
2. JUVOS records from November 1995 back to January 1985 were obtained for people whose most recently recorded address fell within the 80 sampling points used to construct the TfW sample, and who had experienced a spell of unemployment lasting at least five months (regardless of whether this spell had since ended). To ensure confidentiality, the DfEE stripped these records of names, NINOs and the last two postcode digits, and they alone retained the key linking these with the dummy identifiers given to the research team.<sup>2</sup>
3. Matches were chosen for each member of the TfW sample, using three inviolable and two flexible criteria. The inviolable criteria were sex, sampling point, and being unemployed in the month that the matched TfW participant entered TfW (note that although all TfW sample members left TfW at around the same time, there was wide variation in entry dates). The term 'qualifying spell of unemployment' when applied to the comparison sample refers to the spell of unemployment that included this month. The flexible matching criteria were the date at which the qualifying spell of unemployment began, and age. On the first pass through the data, the closest match was found for each TfW participant, making a random choice in the case of ties. Matched TfW participants and JUVOS members were then withdrawn from the pool, and the procedure repeated until all TfW participants had been allocated a match. To allow for sample attrition, the operation was repeated to find the second and third best matches for each TfW participant.
4. Next, the DfEE matched NINOs back in for the selected comparison sample, traced their last known addresses from the database used for the payment of benefits,<sup>3</sup> screened out those who had themselves been on TfW

---

1 Another possible source of information on trainee characteristics was the administrative data recorded on their entry to TfW. However this did not give specific dates for unemployment prior to entry, only length of unemployment in broad bands.

2 Apart from postcodes, addresses are not held on JUVOS. Responsibility for JUVOS has since passed from the DfEE to the Office for National Statistics.

3 Although JUVOS held postcodes, it did not hold addresses.

- (using the database of TfW leavers), and sent out letters giving those selected the opportunity to opt out of the study before an interviewer called at their home. All but the first of these steps entailed some attrition.
5. Addresses were then issued to the field force for interview. It had been planned to issue the second and third best matches only when the better matches were exhausted, but the extent of attrition both at step 4 and in the field meant that all matches had to be used in order to achieve enough interviews. The comparison sample interviews took place at the same time as the Wave 2 TfW sample interviews (June and July 1997), and 929 interviews were achieved. Once again, the lack of accurate and up-to-date addresses was a major source of non-response.

### CLOSENESS OF MATCH: WAVE 1 PARTICIPANT AND SELECTED COMPARISON SAMPLES

For the purposes of the TfW evaluation, the relevant factor in judging the success of the design was not the closeness of the match obtained between the achieved Wave 1 participant sample and the comparison sample initially selected, but the closeness of the match after attrition had taken its toll on both. This is explored in the next section. However since the TfW study there have been significant improvements in the administrative data available for those eligible for government programmes for the unemployed, most notably the New Deal Evaluation Database.<sup>4</sup> With administrative data, sample attrition is minimal, and where it does occur it is much less likely to be correlated with factors relevant to programme outcomes than is the case with interview surveys. Thus the closeness of the match obtained before attrition is of interest, as it indicates the kind of precision achievable in studies based exclusively on administrative data.

On the three inviolable matching criteria (sex, sampling point and being unemployed in the month in which the matched TfW participant entered TfW), the match obtained before attrition was perfect by definition. On the two flexible criteria (start date of the qualifying spell of unemployment and age) the match was very good. Table 1 shows the gap in months between the start date of the TfW participant's qualifying spell of unemployment and that of their best-matched JUVOS partner. As JUVOS began in January 1985, it was impossible to find an exact match for any TfW participant whose

---

<sup>4</sup> This database has been established by a team at the Employment Service under the direction of Mike Daly, using data from a range of administrative sources on people eligible for various New Deal programmes.

**Table 1:** *Closeness of the best match on the start date of the qualifying spell of unemployment, by sex (before attrition)*

	Men %	Women %
exact	66	45
within one month	20	24
within 2–3 months	8	15
within 4–6 months	3	5
within 7–12 months	1	5
more than 12 months	3	6
<b>Total</b>	<b>100</b>	<b>100</b>
<i>N</i>	815	289

qualifying unemployment spell began before then. Nevertheless, two-thirds of male participants were matched exactly and another fifth had a mismatch of just one month. The biggest mismatches were associated with longer unemployment, and so mattered less in proportionate terms. Thus of the 24 men with a mismatch of more than 12 months, nine had begun their qualifying unemployment spell before 1985, and another twelve had begun it before 1990. The match achieved for women participants was not as good as for men, probably because women were heavily outnumbered by men in JUVOS and so there was a much smaller pool from which to seek matches. Nevertheless 69 per cent of women participants were matched either exactly or to within one month, and only 6 per cent had a mismatch of more than 12 months.

Because the second flexible matching criterion, age, was given a lower priority than unemployment start date, the match achieved was not particularly close when assessed on a one-to-one basis, even before attrition. Only about a fifth of participants had a ‘best’ match to within one year of age or less. Nevertheless the mean age of participants was similar to that of the selected comparison sample: 35 years for male TfW participants compared to 36 years for their ‘best’ matches, and 33 years for female participants compared to 35 years for their ‘best’ matches. Overall age distributions were also similar.

These matches were of course found from a JUVOS Cohort that included only 5 per cent of NINOs. With today’s 100 per cent Cohort, we would have even greater success in finding good matches.

## SAMPLE ATTRITION AND RE-MATCHING

The impact of sample attrition on the match achieved is shown by the fact that for 30 per cent of respondents to the Wave 2 interview with TfW participants, we failed to obtain an interview with any of the three people selected as their first, second or third best match. Put the other way around, about a quarter of members of the achieved comparison sample lost their Wave 1 TfW partner through attrition at Wave 2. This figure was similar regardless of whether the comparison sample member was selected as a first, second or third best match.

Despite sample attrition, matching still achieved its main aim of reducing between-sample variation on key confounding variables. We never intended to rely solely on matching to control for all differences between participants and non-participants that could affect outcomes; on the contrary, the surveys collected information on a wide range of potentially relevant factors that were used as control variables in the statistical models of outcomes, alongside the variables used in the matching process. However, for the discrete-time duration models used in the analysis (testing whether TfW participants found jobs more quickly than non-participants) we needed to identify a common starting point, and this we defined as the month in which the qualifying spell of unemployment began. For the TfW sample, this had a clear definition, but for comparison sample members, many of whom had experienced more than one spell of unemployment during the relevant period, it could only be defined by reference to a matched TfW participant. We thus had to conduct a re-matching process.

Re-matching had a second purpose. The initial match had been based on the self-reported unemployment of participants collected at the Wave 1 interview, and on JUVOS data for the comparison sample. However, the analysis had to use the retrospective ten-year work histories collected in the Wave 2 TfW and comparison group interviews: only these could provide a continuous and internally consistent record of activities incorporating both spells of unemployment and time on TfW, and only these were collected in the same way for both samples, using the same definitions and at the same date. Thus we needed to re-match, using interview data rather than JUVOS data for the comparison sample.

The rules for re-matching were less rigorous than the rules for the initial matching process, because we had much smaller numbers to draw from. Members of the comparison sample were excluded if they did not report a spell of unemployment beginning before January 1996 and lasting at least six months. If they had more than one such unemployment spell, we used the spell whose dates matched most closely to the dates of the spell in JUVOS by

virtue of which the person was originally selected. A match was then sought from the Wave 2 TfW sample, according to the following steps, with each step invoked only if more than one TfW sample member remained after the step that preceded it.<sup>5</sup>

1. Choose a TfW participant who entered TfW in a month in which the comparison sample member was unemployed (inviolable condition).
2. Choose the participant whose unemployment spell before TfW started closest to the start of the comparison sample member's unemployment spell.
3. Choose the participant of the same sex as the comparison sample member (if no-one, skip to step 4).
4. Choose the participant closest in age to the comparison sample member.
5. Choose the participant who had previously been chosen least often.
6. If there is still more than one TfW sample member remaining, choose randomly.

## EVALUATION OF THE RE-MATCHED SAMPLES

Matches were found for 815 of the 929 members of the achieved comparison sample and the remaining 114 were excluded from any further analyses. The reasons why no match was found for them were as follows:

- Substantial missing information in the ten-year retrospective work history 16
- No spell of unemployment at all recorded in the work history 15
- No spell of unemployment that began before 1996 6
- No spell of unemployment beginning before 1996 that lasted 6 months or more 53
- Failed the inviolable Step 1 listed in Section 5 24

Many of these problems arose because of differences between the JUVOS record of formal benefit claims and self-reports of claimant unemployment collected at interview.

For the 815 for whom matches were found – henceforth called the ‘re-matched comparison sample’ – the match with their selected TfW partner was very close. There was an exact match on the start date of the qualifying

---

<sup>5</sup> In the re-matching process, a single Wave 2 TfW respondent was allowed to act as the match for more than one member of the achieved comparison sample.

spell of unemployment for 90 per cent, a further 5 per cent were matched to within one month, and all but 0.5 per cent matched to within six months. However, these figures are good because we allowed a single TfW sample member to act as partner for more than one member of the comparison sample. Some members of the re-matched TfW sample were partnered several times, and some were not partnered at all. A more realistic assessment of the comparability of the re-matched samples is obtained from the overall distributions of relevant variables, in which members of the re-matched TfW sample appear once only. Assessed in this way, the re-matched samples were very well matched on sex, with 73 per cent male in the TfW sample and 74 per cent male in the comparison sample. The age match was also reasonably satisfactory (Figure 1). Finally, Figure 2 shows that the re-matched samples had very similar distributions for the start date of the qualifying spell of unemployment, the only significant discrepancy being that the TfW sample had an excess number of spells starting in the first quarter of 1993.<sup>6</sup>

## RESULTS

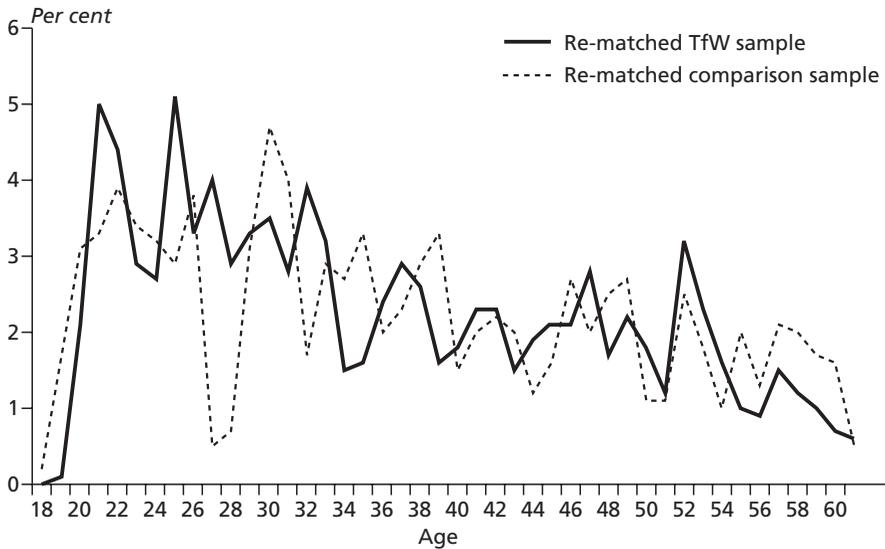
There is space in this paper only for the briefest outline of findings, which are fully reported in Payne, Payne, Lissenburgh and Range (1999). We concentrate on the results that are most relevant to the methodological focus of the paper.

We found that, even after matching, TfW participants were more likely than members of the comparison sample to have also more likely to have worked in clerical or secretarial occupations and to have previously been on government programmes. People were also less likely to go on TfW if they were women with children under school age, had long-term health problems or disabilities, lived in rented accommodation, had poor academic qualifications, were of Indian Subcontinent origin, had no driving licence and no regular access to a vehicle, had never held a job, or already had good vocational qualifications. It is likely that selectivity in entry to TfW was encouraged by TfW's 'starts and outcomes' funding structure, which gave incentives to training providers to select those who were most likely to succeed, and by TfW's dependence on the cooperation of employers in providing placements.

After allowing for measured differences between trainees and the comparison sample, the results suggested that TfW had a positive impact on

---

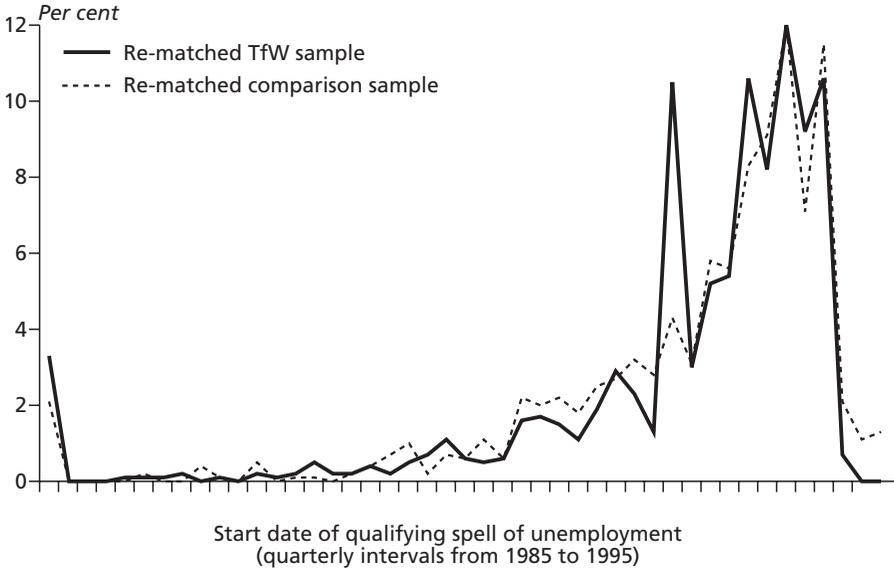
<sup>6</sup> This has been attributed to a change in the rules governing TfW which had the effect of temporarily increasing recruitment in one quarter.



**Figure 1:** *Age: re-matched TfW and comparison samples*

employment prospects. In the year and a half after leaving TfW, participants spent on average around one extra month in seven in work, and this advantage did not appear to fade with time. However this is likely to be a maximal estimate of the impact of TfW, and it is not possible to determine the extent to which the advantage of participants was due to unmeasured differences between them and the comparison sample (Firth, Payne and Payne 1999). More specifically, because participants were first interviewed after leaving TfW rather than on entering the programme, we have no information on possible prior attitudinal and motivational differences between the two samples, which may have increased the chances that participants would find work relative to the comparison sample.

Most of the employment gains associated with TfW came from an increased chance of getting a full-time job as an employee. The impact on rates of entry to part-time or self-employed jobs was comparatively small. Participation did not affect hourly wage rates in subsequent jobs, though it tended to improve take-home pay because it helped more people to secure full-time rather than part-time jobs. Qualifications gained on TfW improved job prospects only if they were of NVQ Level 4 standard or higher. This may have been because trainees preferred to accept a job offer rather than stay on TfW to complete their qualifications, and because the funding structure meant that it was to the advantage of programme providers to encourage participants to do this.



**Figure 2:** *Start date of qualifying spell of unemployment: re-matched TfW and comparison samples*

Not only was there selectivity in entry to TfW, but there was also selectivity between different types of placement within the programme. Entrants had a better chance of getting an employer placement rather than a project placement or full-time off-the-job training if they were young, female, had good vocational qualifications, held a driving licence, had a partner in work, and were not black. Compared to participants in full-time off-the-job training, those on project placements tended to be older, were more likely to have been manual workers or not to have had a job at all, and were less likely to hold a driving licence or to have a partner in full-time work. Employer placements gave trainees a better chance of securing work than either full-time off-the-job training or project placements, partly because those on employer placements had a good chance of being kept on by their placement employer. Project placements had the smallest impact on job prospects, with slight gains emerging only after a period of time. Again, although these results held true after controlling for measured differences between participants in different strands of TfW, we cannot determine the extent of the contribution of unmeasured difference.

## DISCUSSION AND CONCLUSION

In assessing the impact of TfW on the job chances of those who took part, we faced a dual selection problem – selection both into the programme and within the programme. The matched comparison group design helped to control for the former but was of no relevance to the latter. In neither case could we be sure how much of the difference in outcomes between participants and non-participants, or between participants in different strands of the programme, was due to unmeasured differences between these groups that were present initially. This problem besets most non-experimental evaluations of social programmes. Indeed, Auspos, Riccio and White (1999), in their review of evaluation studies of labour market programmes for young people in the USA and Europe, found that the largely non-experimental European evaluations returned positive findings far more often than comparable studies in the USA, where social experiments were much more likely to be used. They suggest that one reason for this difference in results might be that the evaluation methods used in the USA involve more stringent or conservative tests of programme impacts. All this indicates that if we are to succeed in the goal of evidence-based policy making in the UK, then the ethical and political arguments against random allocation designs must be revisited. Similarly, for the practical obstacles to such designs to be overcome, then the need for sound evaluation must be placed higher in the list of policy makers' priorities.

In the absence of social experiments, more effort needs to be put into understanding the process of selection into the programme under evaluation. This might be done, for example, by direct observation of the selection process, or by a qualitative study of people whose measurable characteristics suggest that they have a high probability of participating in the programme under evaluation, but who do not do so. We also need to examine other evidence of whether the programme is 'adding value' to participants. This evidence could be of several different kinds – for example, participants' views of what the programme did for them, objective measures of changes in skill levels, how any qualifications or other certification obtained on the programme were used in applying for jobs, and the relationship between the specific training content of the programme and the jobs obtained afterwards.

In the study described in this paper, the matched comparison group design aimed to assess the actual impact of TfW on participants and the potential impact of TfW on similar people who did not take part. However the extent of selection into TfW made it difficult to extrapolate from this to the programme's potential impact on all who were eligible but did not take part. In general, if programme providers select those who they feel are most likely to benefit from the programme, then we are likely to over-estimate the

programme's potential impact if it were made more widely available. Alternatively, programme providers (encouraged, for example, by an output-related funding system) may select not those who are most likely to gain from the programme, but those who are most likely to achieve a successful outcome, regardless of whether or not they go on the programme. In this case, we may under-estimate the programme's potential impact on a wider population. Again, in order to draw sound conclusions, we need either a random allocation design or a better understanding of the selection process – or, ideally, both.

Recent improvements in the availability of administrative data in the UK (notably the 100 per cent JUVOS cohort and the New Deal Evaluation Database) have greatly increased the options for the design of programme evaluations, even if random allocation designs continue to be ruled out. With good administrative data for participants and eligible non-participants, we have much greater flexibility in how we select a comparison group – or more than one, if appropriate – and it becomes viable to use propensity score matching, with all its advantages. Good administrative data also avoid the problem of sample attrition, which is becoming increasingly serious as response rates to social surveys in the UK continue to fall. Furthermore, they can give us sample numbers that are big enough to test for differential programme effects on comparatively small sub-groups of participants such as ethnic minorities or the over-50s. Administrative data can also relieve the timetable pressures that have impeded much UK evaluation research over the last decade, as they allow us to derive quick answers from archived data. Similarly, they make it possible to get information on the characteristics of programme participants before the process of selection into the programme begins, to chart their experiences while on the programme, to track programme drop-out, and to follow programme leavers over a long enough period to judge the stability of programme outcomes. Probably the single most effective way of improving the quality of evaluation research in the UK is to continue to enhance administrative databases. If this were combined with a more permissive approach towards social experiments, then the outlook for evidence-based policy making would be rosy.

## REFERENCES

- Auspos, P, J Riccio, and M White (1999) *A Review of the US and European Literature on the Microeconomic Effects of Labour Market Programmes for Young People*. Sheffield: Employment Service Research and Development Report ESR20

- Burtless, G and L Orr (1986) 'Are classical experiments needed for manpower policy?'. *Journal of Human Resources* 21, 606–39
- Department for Education and Training (1997) 'Government-Supported Training: YT/MA/TfW England and Wales'. London: DfEE Press Notice 263/97
- Firth, D, C Payne and J Payne (1999) 'Efficacy of programmes for the unemployed: discrete-time modelling of duration data from a matched comparison study'. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 162, 111–120
- Heckman, J, R LaLonde and J Smith (1999) 'The economics and econometrics of active labor market programs'. In *Handbook of Labor Economics* Vol III (ed O Ashenfelter and D Card)
- Lalonde, R (1986) 'Evaluating the econometric evaluations of training programs with experimental data'. *American Economic Review* 76, 604–20
- Manski, C and I Garfinkel (1992) 'Introduction' In *Evaluating Welfare and Training Programs* (ed C Manski and I Garfinkel) Cambridge: Harvard University Press
- Payne, J, C Payne, S Lissenburgh and M Range (1999) *Work-Based Training and Job Prospects for the Unemployed: An Evaluation of Training for Work*. Sheffield: Department for Education and Employment Research Report RR96
- Sproston, K (1999) *TfW: Survey of Ex-Participants – A Report on Phases 1 and 2*. London: Social and Community Planning Research (SCPR Project 1544/1657)
- White, M and J Lakey (1992) *The Restart Effect: Does active labour market policy reduce unemployment?*. London: PSI Publishing